確率・統計基礎: ベイズ推定

森 立平

用語

- $x \in A$: データ、知っているもの (今日は離散と仮定するが、本質的な仮定ではない)
- $s \in B$: パラメータ、推定したいもの (今日は離散の場合を考える。連続のときは θ という記号を使うことが多い)
- $\Pr(S = s)$: 事前確率
- $\Pr(X = x \mid S = s)$: 尤度
- $\Pr(S = s \mid X = x)$: 事後確率

最大事後確率推定

問題

確率分布 $\Pr(S = s, X = x)$ が既知である。与えられた $x \in A$ から $s \in B$ を推定したい。推定関数 $E\colon A \to B$ の平均誤り確率は $\Pr(E(X) \neq S)$ で定義される。

$$
\begin{aligned}
\Pr(E(X) \neq S) &= 1 - \Pr(S = E(X)) \\
&= 1 - \sum_{x \in A} \Pr(X = x) \Pr(S = E(X) \mid X = x) \\
&= 1 - \sum_{x \in A} \Pr(X = x) \Pr(S = E(x) \mid X = x)
\end{aligned}
$$

これを最小化したいので $E(x) = \arg\max_s \Pr(S = s \mid X = x)$ とするのが最適である。これを最大事後確率推定という。

最尤推定

最大事後確率推定 (Maximum a posteriori probability estimation)

$$E_{\mathrm{MAP}}(x) \coloneqq \operatorname*{arg\,max}_{s} \Pr(S = s \mid X = x)$$

平均誤り確率を最小化する。事前分布を知らないと使えない。

最尤推定 (Maximum likelihood estimation)

$$E_{\mathrm{ML}}(x) \coloneqq \operatorname*{arg\,max}_{s} \Pr(X = x \mid S = s)$$

平均誤り確率については何も言えない。事前分布を知らなくても使える。

事前分布が一様分布 $P(S = s) = \frac{1}{|B|}$ のとき、

$$\begin{aligned}
\operatorname*{arg\,max}_{s} \Pr(S = s \mid X = x) &= \operatorname*{arg\,max}_{s} \frac{\Pr(S = s, X = x)}{\Pr(X = x)} \\
&= \operatorname*{arg\,max}_{s} \Pr(S = s, X = x) \\
&= \operatorname*{arg\,max}_{s} \Pr(S = s) \Pr(X = x \mid S = s) \\
&= \operatorname*{arg\,max}_{s} \Pr(X = x \mid S = s)
\end{aligned}$$

最大事後確率推定と最尤推定は等しい。

二値の推定

$B = \{0, 1\}$ の状況を考える。以下のように $p_0(x), p_1(x)$ を定義する。

$$
\begin{aligned}
p_0(x) &\coloneqq \Pr(X = x \mid S = 0) \\
p_1(x) &\coloneqq \Pr(X = x \mid S = 1)
\end{aligned}
$$

$\Pr(S = 0) = \lambda$, $\Pr(S = 1) = 1 - \lambda$ とする。最大事後確率推定では $\lambda p_0(x)$ と $(1 - \lambda) p_1(x)$ を比較して、前者 (もしくは後者) が大きい場合 0(もしくは 1) と推定する。このとき

$$
\begin{aligned}
&\Pr(E(X) = S) - \Pr(E(X) \neq S) \\
&= \sum_s \Pr(S = s) \Pr(E(X) = S \mid S = s) - \sum_s \Pr(S = s) \Pr(E(X) \neq S \mid S = s) \\
&= \sum_{x \in E^{-1}(0)} \lambda p_0(x) + \sum_{x \in E^{-1}(1)} (1 - \lambda) p_1(x) \\
&\quad - \left(\sum_{x \in E^{-1}(1)} \lambda p_0(x) + \sum_{x \in E^{-1}(0)} (1 - \lambda) p_1(x) \right) \\
&= \sum_{x \in E^{-1}(0)} (\lambda p_0(x) - (1 - \lambda) p_1(x)) - \left(\sum_{x \in E^{-1}(1)} (\lambda p_0(x) - (1 - \lambda) p_1(x)) \right) \\
&= \sum_{x \in A} |\lambda p_0(x) - (1 - \lambda) p_1(x)|
\end{aligned}
$$

全変動距離

最小誤り確率は

$$\frac{1 - \sum_{x \in A} |\lambda p_0(x) - (1-\lambda) p_1(x)|}{2}.$$

特に $\lambda = 1/2$ のとき、

$$\frac{1 - \frac{1}{2} \sum_{x \in A} |p_0(x) - p_1(x)|}{2}.$$

Definition (全変動距離 (Total variation distance))

$p, q \in \mathcal{P}$ についてその全変動距離 $d_{\mathrm{TV}}(p, q)$ を以下で定義する。

$$d_{\mathrm{TV}}(p, q) \coloneqq \frac{1}{2} \sum_{x \in A} |p(x) - q(x)|.$$

平均誤り確率の漸近的挙動

二種類あるコインの片方を選んで独立に何回も投げる。どちらのコインを選んだのか推定したい。

$$\Pr(\mathbf{X} = \mathbf{x},\, S = s) = \Pr(S = s) \prod_{k=1}^{N} \Pr(X_i = x_i \mid S = s)$$

誤り確率は

$$\sum_{\mathbf{x} \in A^N} \left| \lambda \prod_k p_0(x_k) - (1-\lambda) \prod_k p_1(x_k) \right|.$$

から定まるが、これを見ても $N \to \infty$ の挙動はよく分からない。
$p(s) := \Pr(S = s),\ p(x \mid s) := \Pr(X = x \mid S = s),\ p(\mathbf{x} \mid s) := \prod_k p(x_k \mid s)$
とすると、

$$\begin{aligned}\Pr\left(E(\mathbf{X}) \neq S\right) &= \Pr\left(p(S)p(\mathbf{X} \mid S) \le p(\bar{S})p(\mathbf{X} \mid \bar{S})\right) \\ &= \Pr\left(\sum_k \log \frac{p(X_k \mid S)}{p(X_k \mid \bar{S})} \le \log \frac{p(\bar{S})}{p(S)}\right).\end{aligned}$$

(不等号が等号の場合は 1/2 の確率でエラーになるが、簡単のためそれはエラーと数えることにする)。

平均誤り確率の漸近的挙動

$$
\begin{aligned}
\Pr\left(\sum_k \log \frac{p(X_k \mid S)}{p(X_k \mid \overline{S})} \le \log \frac{p(\overline{S})}{p(S)}\right) &= \lambda \Pr\left(\sum_k \log \frac{p(X_k \mid 0)}{p(X_k \mid 1)} \le \log \frac{1-\lambda}{\lambda} \mid S = 0\right) \\
&\quad + (1-\lambda) \Pr\left(\sum_k \log \frac{p(X_k \mid 1)}{p(X_k \mid 0)} \le \log \frac{\lambda}{1-\lambda} \mid S = 1\right)
\end{aligned}
$$

この確率はクラメールの定理で漸近評価できる。

課題

$A = \{0, 1, 2\}$, $B = \{0, 1\}$ とし、

$$
\begin{array}{ccc}
\Pr(S = 0) = 1/4 & \Pr(S = 1) = 3/4 & \\
\Pr(X = 0 \mid S = 0) = 1/6 & \Pr(X = 1 \mid S = 0) = 2/6 & \Pr(X = 2 \mid S = 0) = 3/6 \\
\Pr(X = 0 \mid S = 1) = 3/6 & \Pr(X = 1 \mid S = 1) = 1/6 & \Pr(X = 2 \mid S = 1) = 2/6
\end{array}
$$

とするとき、以下の問に答えよ。

1. X から S を最大事後確率推定する関数 $E_{\mathrm{MAP}} : A \to B$ と最尤推定する関数 $E_{\mathrm{ML}} : A \to B$ をもとめよ。
2. X から S を最大事後確率推定した場合の平均誤り確率をもとめよ。
3. X から S を最尤推定した場合の平均誤り確率をもとめよ。